
Un modèle de graphe et de dioïde pour le routage interdomaine

Marc-Olivier Buob, Mickaël Meulle, Jean-Luc Lutton

France Télécom R&D
38-40, rue du Général Leclerc
92794 Issy les Moulineaux Cedex 9
{marcolivier.buob,michael.meulle,jeanluc.lutton}@francetelecom.com

RÉSUMÉ. Le routage Internet, assuré par le protocole BGP (Border Gateway Protocol), dépend des choix de chaque opérateur Internet (ou ISP). Pour s'interconnecter, deux ISP négocient un contrat économique. Ces contrats, une fois traduits en règles de routage, impactent considérablement la structure du routage Internet (routage interdomaine). En effet, le routage interdomaine n'est pas un routage aux plus courts chemins, puisque les choix de routage BGP sont soumis aux politiques de routage inconnues des ISP. Il est alors difficile de prévoir et inférer les cheminements Internet de bout en bout. Pourtant cette information est importante pour un opérateur, notamment pour appliquer des techniques de Traffic Engineering ou pour évaluer l'impact d'un changement de politique de routage. Dans cet article, nous proposons un modèle de cheminement Internet à la granularité AS avec prise compte des annonces sélectives sur les préfixes réseau. Ce modèle, consistant en une transformation de graphe et une structure algébrique de dioïde, permet d'inférer les routages interdomaines sans les simuler. Les routes calculées depuis chaque AS vers un préfixe donné, respectent les accords économiques et les deux premières règles du processus de décision BGP simplifiées.

ABSTRACT. Business relationships between Internet Service Providers (ISPs), which remain confidential, have a huge impact on the structure and behaviour of Internet routing. Those agreements are a key point in policy-based Internet routing and thus in determining ISP ranking and hierarchical structure and in interdomain traffic engineering methods. In this article, we propose a model of Internet routing at the Autonomous System (AS) level, with per-prefix selective announcements. We simulate interdomain route computation without simulation of the interdomain routing protocol BGP. Though we are able to study end-end IP traffic flows, and what-if scenarios for BGP routing policy changes. Our model computes routes from each AS to each IP network that respect rules of business agreements and an approximation of the first two rules of the BGP best route decision process. We finally show the effectiveness of Dijkstra algorithm with our graph transformation and the algebraic structure of interdomain link-weights we defined.

MOTS-CLÉS : Internet, BGP, routage interdomaine, modèle de graphe, dioïde

KEYWORDS: Internet, BGP, interdomain routing, graph models, path algebra

1. Introduction

L'interconnexion entre deux opérateurs Internet (ou ISP¹) s'effectue en négociant un contrat bilatéral généralement de type client / fournisseur² ou peering³. Selon la nature des contrats négociés avec les autres opérateurs, un ISP va définir des règles de préférence et de retransmission des routes Internet dont il dispose. Ces règles et préférences sont retranscrites dans le réseau d'un ISP au travers des configurations de ses routeurs. En effet, le protocole interdomaine BGP⁴ utilisé par tous les opérateurs Internet, permet la définition de telles politiques de routage. Chaque opérateur délimite son réseau en un ou plusieurs systèmes autonomes (AS) et lui associe une politique de routage BGP unique et cohérente. Le routage interdomaine BGP permet à chaque routeur la constitution d'une table de routes qui ont pour destination des plages d'adresses IP publiques (préfixes). Ce routage n'est pas un routage au plus court chemin, car il découle des politiques de routage de chaque AS.

Il est difficile de prévoir les routes BGP sélectionnées par les routeurs de chaque AS, puisque les politiques de routage sont inconnues et influencent la préférence et la propagation des routes BGP. Pourtant, la connaissance des routes interdomaines de bout en bout est un point clé dans les études de Traffic Engineering [QUO 05].

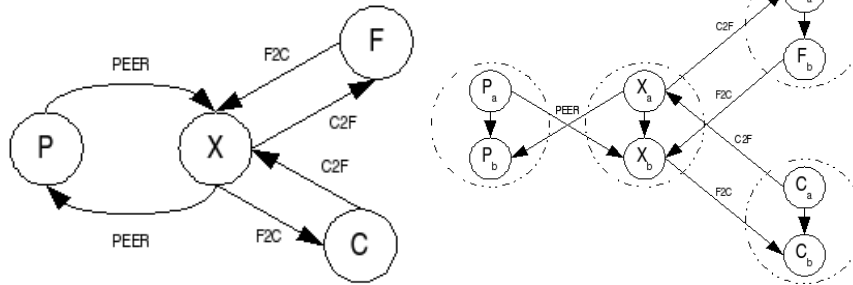
Dans cet article, nous proposons une méthodologie permettant le calcul de routes BGP qui respectent les règles économiques entre AS, et qui tient compte des annonces sélectives de routes intervenant dans le cas d'AS multihomés⁵ [WAN 03]. L'intérêt de notre approche est d'éviter la simulation exhaustive des échanges des messages BGP [QUO 05] en dépit d'une perte de précision sur les chemins inférés. Dans un premier temps nous présenterons le modèle topologique de graphe utilisé pour représenter l'Internet à la granularité AS en tenant compte des différents réseaux annoncés. Dans un second temps, nous détaillerons la structure algébrique des poids associés aux chemins du graphe, qui nous permet d'appliquer l'algorithme de Dijkstra généralisé [M.G 02] [M.M 04] pour le calcul des routes BGP.

2. Un modèle topologique de graphe AS-préfixe

2.1. Le graphe *inter-AS*

L'Internet interdomaine peut être représenté par un graphe de sessions BGP entre routeurs. On s'intéresse ici à une vision moins fine, où l'on associe à chaque AS un unique sommet du graphe. Lorsque deux systèmes autonomes sont connectés par le biais de sessions BGP entre leurs routeurs, on ajoute seulement deux arcs orientés

-
1. Internet Service Provider.
 2. le client paye pour obtenir les routes Internet de son fournisseur et lui annoncer les routes vers ses clients.
 3. les deux opérateurs échangent gratuitement les routes Internet à destination de leurs clients.
 4. Border Gateway Protocol
 5. Un AS est dit multihomé lorsqu'il à plusieurs fournisseurs.



(a) Un exemple de graphe inter-AS : C est client de X , P et X sont en peering, et F est un fournisseur de X .

(b) Graphe inter-AS étendu associé

reliant les deux sommets du graphe. On associe à chaque arc un unique accord d'interconnexion représenté par une étiquette. On considère en effet que les opérateurs ne font intervenir qu'un seul type de contrat entre deux AS⁶. Les différents contrats considérés sont :

C2F accord d'un sommet client vers un sommet fournisseur,

F2C accord d'un sommet fournisseur vers un sommet client,

PEER accord entre deux peers.

On obtient alors un graphe $G_{as}(V_{as}, E_{as})$ dont un exemple est donné figure 1(a).

La nature des contrats établis entre les AS implique quelques règles simples concernant la propagation des routes BGP. En effet, si un AS propage ses routes provenant d'un AS fournisseur (ou d'un AS peer), à un autre AS fournisseur (ou peer), alors il sera susceptible de véhiculer un trafic qui chargera son réseau tout en lui coûtant de l'argent. En conséquence, les chemins d'AS propagés par les routeurs respectent le motif suivant⁷ :

$$(C2F)^*(PEER)?(F2C)^*$$

Un chemin respectant un tel motif est donc constitué d'une suite (éventuellement vide) d'arcs $C2F$, suivi (éventuellement) d'un arc $PEER$, et enfin d'une suite (éventuellement vide) d'arcs $F2C$. Ces chemins d'AS sont dits *économiquement valides* ou *valley-free* (cf. [GAO 01] pour plus de détails).

On peut remarquer qu'un chemin d'AS est économiquement valide si et seulement si chacune des paires d'arcs successives qui le composent sont *économiquement valides* (par exemple le chemin $ABCDE$ est *économiquement valide* si et seulement si les paires (A, B) , (B, C) , (C, D) , (D, E) le sont) :

- un arc $C2F$ peut être suivi d'un arc $C2F$, $PEER$, ou $F2C$,

6. En réalité, certains AS peuvent convenir à des contrats hybrides transit/peering.

7. sauf erreur de configuration.

- un arc *PEER* peut être suivi d'un arc *F2C*, mais pas d'un arc *PEER* ou *C2F*
- un arc *F2C* peut être suivi d'un arc *F2C*, mais pas d'un arc *PEER* ou *C2F*.

De ces paires dites *économiquement valide* va découler la transformation de graphe présentée dans la section 2.2.

2.2. Le graphe inter-AS étendu

Nous cherchons à calculer des chemins *économiquement valides* dans le graphe inter-AS. Pour cela, nous allons modifier la structure du graphe de sorte à garantir la validité économique de *tous* les chemins d'AS que l'on peut construire. On pourra donc grâce à cette transformation de graphe s'affranchir de la notion d'accord pour l'algorithme de cheminement présenté dans la section 3. Chaque sommet correspondant à un AS X est décomposé en deux sommets X_a et X_b de la façon suivante :

- Pour chaque AS C (représenté par les sommets C_a et C_b) client de X , on construit l'arc (C_a, X_a) avec l'étiquette *C2F*, et l'arc (X_b, C_b) avec l'étiquette *F2C*.
- Pour chaque AS P (représenté par les sommets P_a et P_b) en peering avec l'AS X , on construit les arcs (P_a, X_b) et (X_a, P_b) avec l'étiquette *PEER*.
- Pour chaque AS F (représenté par les sommets F_a et F_b) fournisseur de X , on construit l'arc (F_b, X_b) avec l'étiquette *F2C*, et l'arc (X_a, F_a) avec l'étiquette *C2F*.

On note $G_e(V_e, E_e)$ le graphe inter-AS étendu ainsi formé. On pourra se reporter à la figure 1(b), qui représente le graphe de la figure 1(a) après transformation. Un chemin économiquement valide dans le graphe inter-AS entre deux sommets X et Y est alors équivalent à un chemin dans le graphe inter-AS étendu entre les sommets X_a et Y_b .

Soient V_e^{as} l'ensemble des sommets représentant un AS dans G_e , et E_e^{as} l'ensemble des arcs inter-AS du graphe G_e . On a alors :

$$(|V_e^{as}|, |E_e^{as}|) = (2 \times |V_{as}|, |E_{as}| + |V_{as}|)$$

Le modèle de graphe inter-AS ne tient pas compte des différents préfixes annoncés par chaque AS. Pourtant, chaque AS a la possibilité de choisir les préfixes qu'il annonce à chacun de ses voisins. C'est le cas notamment des AS multihomés. Nous allons, pour répondre à cette problématique, compléter ce modèle de graphe étendu afin de prendre en compte la notion d'annonces sélectives.

2.3. Un modèle de graphe AS-préfixe étendu

Le modélisation de l'Internet interdomaine s'appuie notamment sur deux hypothèses simplificatrices. D'une part, on suppose que deux AS partenaires établissent le même contrat d'interconnexion sur chacun de leurs liens. Cette hypothèse ne traduit

pas tout à fait la réalité car selon la localisation géographique, l'accord d'interconnexion peut différer. D'autre part, on ne distingue pas en terme de graphe un AS de ses préfixes. On ne peut donc pas tenir compte de politiques d'annonces sélectives basées sur les préfixes, comme dans le cas des AS multihomé. Par exemple, un AS peut décider de n'exporter que ses préfixes européens à ses partenaires européens et que ses préfixes américains à ses partenaires américains. C'est ce second problème que nous allons tenter de résoudre, en introduisant dans notre modèle de graphe la notion de préfixe.

On ajoute aux graphes G_e un sommet par préfixe. Comme le nombre de préfixes est très important, nous allons ajouter des sommets virtuels (appelés par la suite *agrégats de préfixes*) permettant de diminuer le nombre d'arcs dans le graphe. Soit un AS X annonçant un sous-ensemble de ses préfixes à p_1, p_2, \dots à un de ses AS partenaire Y . On ajoute au graphe le sommet virtuel A_X^Y , représentant l'*agrégat de préfixes* que X transmet à Y , tel que :

- le sommet A_X^Y est relié à chaque préfixe que X annonce à Y par un arc portant l'étiquette *INC*,
- Un sommet X_c associé à l'AS X est relié à l'*agrégat de préfixe* A_X^Y par un arc muni de l'étiquette *OWN*. On, relie aussi le nouveau sommet X_c au sommet X_a pour garantir l'accès de l'AS X à ses préfixes,
- l'AS Y est relié au sommet A_X^Y par l'intermédiaire d'un arc entre le sommet Y_a (respectivement Y_b) et le sommet A_X^Y portant le même accord que l'arc orienté (Y, X) si l'AS X est client (respectivement peer ou fournisseur) de l'AS Y .

On note $G_{pe}(V_{pe}, E_{pe})$ le graphe ainsi obtenu par transformation de G_e . Dans ce graphe, chaque AS est représenté par trois noeuds virtuels. Les autres sommets représentent les *agrégats de préfixes*, et les préfixes. Un chemin économiquement valide entre un AS X et un préfixe p est alors équivalent à un chemin dans le graphe AS-préfixe étendu entre les sommets X_c et p .

La figure 1 montre un exemple de graphe AS-préfixes étendu où l'AS X annonce les préfixes p_1, p_2, p_3 à l'AS P (agrégat A_X^P) et les préfixes p_1, p_4 à l'AS C (agrégat A_X^C). Pour alléger la figure, on n'a pas représenté d'agrégat A_X^F représentant les préfixes que X envoie à F .

Le graphe AS-préfixe étendu permet donc d'avoir un modèle topologique pour la formation de chemins économiquement valides entre des AS et des préfixes. Dans ce graphe on ne calcule plus des chemins d'AS, mais des routes BGP, i.e. des chemins d'AS à destination d'un préfixe. On est donc en mesure d'observer des chemins différents depuis un même AS source vers deux préfixes d'un AS multihomé. Le graphe AS-préfixe étendu, sensiblement plus grand que le graphe inter-AS, reste de taille suffisamment raisonnable pour permettre l'utilisation de l'algorithme de cheminement présenté dans la section 3.

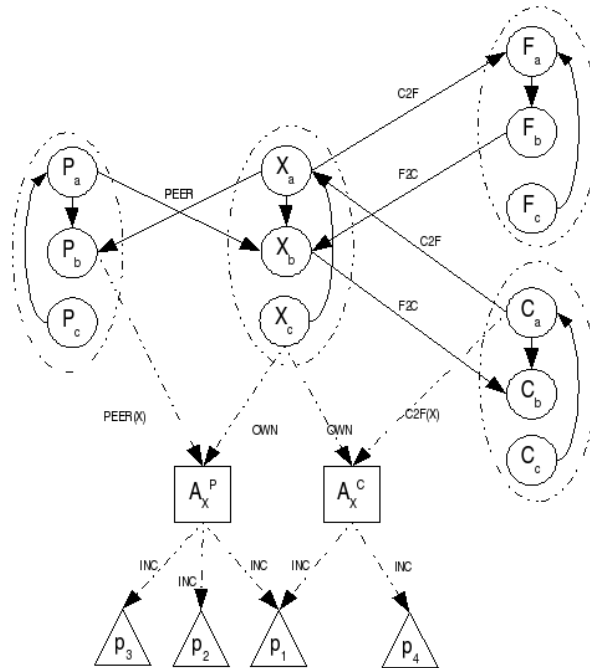


Figure 1 – Un exemple de graphe AS-préfixes étendu (les sommets virtuels associés aux AS sont symbolisés par des cercles, les préfixes par des triangles, et les agrégats de préfixes par des carrés)

3. Modèle de cheminement

On cherche à trouver un modèle de poids pour les arcs permettant de calculer des chemins d'AS à destination d'un préfixe (routes BGP) vérifiant les propriétés suivantes :

- le chemin est *économiquement valide*,
- le chemin est le meilleur au sens du processus de décision BGP.

La première propriété étant assurée par le modèle de graphe étendu (voir paragraphe 2.3), nous allons maintenant mettre en place un modèle pour chercher les meilleurs chemins au sens BGP. Nous allons d'abord rappeler brièvement le processus de décision BGP. Nous présenterons ensuite l'algorithme et la structure algébrique utilisés.

3.1. Le processus de décision BGP simplifié

Un routeur alimente ses tables de routage grâce aux protocoles qu'il implémente

(BGP⁸, OSPF⁹, ISIS¹⁰, ...) et avec les routes statiques configurées manuellement. Lorsqu'un routeur reçoit une annonce BGP et qu'il décide de l'accepter (*import policy*), il l'intègre dans une table appelée Adj-RIB-in. Il compare ensuite les différentes routes relatives au préfixe reçu (*BGP decision process*) et installe la meilleure route dans sa table de forwarding (FIB). La meilleure route est ensuite modifiée et envoyée vers les routeurs autorisés par la politique BGP d'export. Afin de sélectionner la meilleure route, les routeurs BGP comparent successivement les attributs BGP des différentes routes connues pour chaque préfixe jusqu'à un tie break éventuel. Le processus de décision BGP (simplifié) consiste à :

- 1) Préférer la route avec le plus grand *LOCAL_PREF*. Cet attribut est en général associé à un *AS*, notamment pour caractériser l'accord économique passé entre les deux *AS*.
- 2) Préférer la route empruntant le chemin d'*AS* (*AS_PATH*) le plus court.
- 3) Préférer les routes apprises par IGP (ISIS, OSPF...) à celles apprises par BGP
- 4) Si le next-hop *AS* est le même pour les deux annonces, préférer la route avec le plus petit *MED*¹¹. Le *MED* sert à privilégier un point de sortie par rapport à un autre.
- 5) Préférer les routes apprises par *eBGP* (apprise par un routeur BGP du même *AS*) à celles apprises par *iBGP*.
- 6) Préférer l'annonce ayant le plus petit poids IGP
- 7) ...

Un modèle de graphe à la granularité *AS* et préfixe pour le routage interdomaine ne peut prendre en compte que des attributs de routage BGP indépendants des différents routeurs et du routage interne de chaque *AS*. Dans la pratique, les différents attributs BGP sont fixés et modifiables par chaque routeur et le poids IGP vers chaque point de sortie BGP change de routeur à routeur. Puisque la granularité de notre modèle topologique de graphe ne pourra capturer que des chemins d'un *AS* en destination d'un autre *AS* ou d'un préfixe, le choix simplifié BGP d'un *AS* se limite au choix du prochain *AS* à traverser. Le processus de décision d'un *AS* dans notre modèle correspondrait à une décision mutuelle de tous les routeurs d'un même *AS*. Ce cas de diversité de route se produit lorsqu'il existe un unique *AS-PATH* gagnant dans une des deux premières étapes du processus de décision BGP (ce qui est fréquent pour une majorité d'*AS*). Les décisions de routage effectuées au delà de la deuxième étape du processus de décision BGP (qui ne peuvent pas être distinguées) seront considérées comme des choix de routage équivalents. Les attributs correspondant à des considérations intra-*AS*, (valeurs de *MED*, *LOCAL_PREF* associés à un point d'entrée ou de sortie d'un *AS*), et l'attribut *community* ne seront pas pris en compte. Finalement, les attributs *LOCAL_PREF* et *AS_PATH* sont les seuls à considérer dans un pro-

8. Border Gateway Protocol.

9. Open Shortest Path First.

10. Intermediate System - Intermediate System.

11. Multi Exit Discriminator.

cessus de décision BGP simplifié pour la granularité AS. Les valeurs des attributs *LOCAL_PREF* utilisées par chaque AS sont inconnues et ne sont pas toujours relatives uniquement à l'AS voisin. Cependant, les règles usuelles d'utilisation indiquent que les routes BGP apprises grâce aux clients sont préférées aux routes apprises grâce aux peers, elles-mêmes préférées aux routes apprises grâce aux fournisseurs¹². Même si les accords d'interconnexion restent inconnus et que le nombre de solutions à ce problème est exponentiel, il est envisageable d'obtenir une estimation de ces valeurs grâce aux méthodes issues de la littérature. Puis nous définissons les *LOCAL_PREF* (LP) suivants : $LP(F2C) > LP(PEER) > LP(C2F)$.

A notre niveau de granularité on ne considère qu'un seul lien entre deux AS donnés. En effet l'inférence des accords d'interconnexion dont découle l'affectation des *LOCAL_PREF* ne donne qu'un seul accord pour chaque paire d'AS. Une des pistes en cours d'exploration consiste à réajuster ces *LOCAL_PREF* en vue de faire coïncider les chemins calculés avec les chemins mesurés.

Notre modèle de poids de répétition permet de prendre en compte le phénomène d'AS prepending pour les annonces de préfixes en associant un poids à chaque arc *INC* d'un agrégat vers un noeud préfixe¹³. En revanche pour les répétitions appliquées sur les routes transitant entre AS, on ne peut prendre qu'un seul type de répétition par préfixe transmis par un AS voisin. Nous avons choisi d'utiliser les valeurs de répétitions minimales observées.

3.2. Structure algébrique utilisée

Introduction :

Minoux [M.G 02] propose une généralisation de l'algorithme de Dijkstra pour des métriques plus générales que des valeurs numériques. Cependant, l'espace de métriques associées aux chemins, noté \mathcal{E} , doit vérifier certaines hypothèses. On munit \mathcal{E} de deux lois :

- Une loi de comparaison (notée par la suite \oplus) : $a \oplus b = a$ si le chemin a est plus court que le chemin b .
- Une loi de concaténation (notée par la suite \otimes) : $a \otimes b$ est la métrique du chemin formé successivement de a et b .

Définition :

Soit un ensemble \mathcal{E} muni des lois internes \oplus et \otimes . On dit que $(\mathcal{E}, \oplus, \otimes)$ est un *dioid* [M.G 02] si et seulement si :

12. En effet, un opérateur a intérêt pour joindre un préfixe, lorsque c'est possible, à passer par un client (qui le paiera), sinon par un peer, et sinon par un fournisseur (qu'il devra payer). On supposera qu'un AS accorde autant d'importance à chacun de ses clients (resp. à chacun de ses peers, resp. à chacun de ses fournisseurs).

13. Les routeurs AS sont en effet susceptibles de répéter plusieurs fois leur ASN dans les attributs *AS_PATH* des routes BGP qu'ils annoncent via eBGP.

- (\mathcal{E}, \oplus) est un monoïde commutatif (avec ε neutre de \oplus),
- (\mathcal{E}, \otimes) est un monoïde (avec e neutre de \otimes),
- la relation de préordre canonique est une relation d'ordre, (i.e. si la relation d'ordre \leq_c définie par $\forall x, y \in \mathcal{E}, x \leq_c y \Leftrightarrow \exists z \in \mathcal{E}, x \oplus z = y$ est antisymétrique),
- ε est un élément absorbant de \otimes ,
- \otimes est distributive relativement à \oplus .

Si \otimes est seulement distributive à gauche (resp. à droite) relativement à \oplus , on dit que $(\mathcal{E}, \oplus, \otimes)$ est un *diïde à gauche* (resp. à droite).

Modèle utilisé :

Soit $\mathcal{L} = \mathbb{R} \cup \{+\infty\}$ l'ensemble des valeurs que peuvent prendre les *LOCAL_PREF*. Soit $\mathcal{P} = \mathbb{N} \cup \{+\infty\}$ l'ensemble des longueurs possibles des *AS_PATH*. On note $\mathcal{E} = \mathcal{L} \times \mathcal{P}$ l'ensemble des métriques BGP caractérisant les poids associés aux chemins du graphe. On pose : $e = (+\infty, 0)$; $\varepsilon = (0, +\infty)$.

On définit les deux lois de composition interne \oplus et \otimes sur \mathcal{E} par :

$$\forall x_1 = (l_1, s_1), x_2 = (l_2, s_2) \in \mathcal{E}, x_1 \oplus x_2 = \begin{cases} x_1 & \text{si } l_1 > l_2 \text{ ou } (l_1 = l_2 \text{ et } s_1 \leq s_2) \\ x_2 & \text{sinon} \end{cases} \quad (1)$$

$$\forall x_1 = (l_1, s_1), x_2 = (l_2, s_2) \in \mathcal{E}, x_1 \otimes x_2 = \begin{cases} (l_1, s_1 + s_2) & \text{si } x_2 \neq \varepsilon \\ \varepsilon & \text{sinon.} \end{cases} \quad (2)$$

La loi \oplus caractérise dans notre cas, les deux premières étapes du processus de décision BGP (voir paragraphe 3.1). La loi \otimes permet de caractériser la propagation d'un message BGP.

On vérifie que $(\mathcal{E}, \oplus, \otimes)$ n'est pas un diïde car l'élément e n'est pas neutre à gauche pour \otimes . Mais la distributivité à gauche va permettre l'application de l'algorithme de Dijkstra généralisé transposé dans le sens n sources, 1 destination. La non-distributivité à droite peut s'illustrer par le contre-exemple suivant :

$$a = (1, 1); b = (2, 2); c = (3, 3)$$

$$a \otimes (b \oplus c) = (1, 1) \otimes ((2, 2) \oplus (3, 3)) = (1, 1) \otimes (3, 3) = (1, 4)$$

$$(a \otimes b) \oplus (a \otimes c) = ((1, 1) \otimes (2, 2)) \oplus ((1, 1) \otimes (3, 3)) = (1, 3) \oplus (1, 4) = (1, 3)$$

La structure algébrique distributive à gauche, sélective pour \oplus avec e absorbant et plus grand élément, et la non-existence de circuit 0-absorbant sont suffisantes pour que l'on puisse appliquer l'algorithme de Dijkstra généralisé. La preuve est détaillée dans [MEU 07].

L'algorithme de Dijkstra généralisé (modifié) :

Données d'entrée :

$G(V, E)$: le graphe AS-préfixes étendu
 $(\mathcal{E}, \oplus, \otimes)$: le dioïde décrivant les métriques associées aux chemins.
 W : la matrice des poids des arcs du graphe G ($W_{ij} = \varepsilon$ si aucun arc)
 dst : la destination

Données de sortie :

π : le vecteur de poids
 $succ$: le vecteur de successeurs

Algorithme

$\forall i \in V \setminus \{dst\}, \pi(i) \leftarrow \varepsilon, succ(i) \leftarrow i$
 $\pi[dst] = e, succ[dst] = dst$
 $S = \{dst\}$
Tant que $S \neq \emptyset$ faire :
 Soit $i \in S$ le sommet tel que $\pi(i) = \bigoplus_{j \in S} \pi(j)$
 $S \leftarrow S \setminus \{i\}$
Pour chaque $(j, i) \in E$ faire :
 $\pi'(j) \leftarrow \pi(j) \oplus (W(j, i) \otimes \pi(i))$
Si $\pi'(j) \neq \pi(j)$ Alors :
 $\pi(j) = \pi'(j)$ et $succ(j) = i$
 $S \leftarrow S \cup \{j\}$

Cette structure permet de calculer des chemins optimaux au sens BGP, mais ne garantit pas leur validité économique. On pourrait considérer l'espace des métriques $\mathcal{E} \times D \times D$ avec $D = \{C2F, PEER, F2C\}$ et corriger les lois \oplus et \otimes en conséquence. Cependant, la prise en compte des accords d'interconnexion dans l'espace des métriques rompt la distributivité à gauche, et on ne peut dès lors plus appliquer l'algorithme de Dijkstra généralisé sur le graphe AS-préfixes. C'est pourquoi nous avons introduit la notion de graphe AS-préfixe étendu (voir paragraphe 2.3) : les chemins que l'on peut construire dans ce graphe étant économiquement valide, on peut prendre comme espace de métrique l'ensemble \mathcal{E} . Ainsi, le graphe étendu G_{pe} permet de prendre en compte la validité économique des chemins, et l'algorithme de Dijkstra généralisé (appliqué à $(\mathcal{E}, \oplus, \otimes)$) permet de calculer les plus courts chemins au sens BGP. Notons que le modèle de graphe étendu va nécessiter d'appliquer le LOCAL_PREF associé à tous les clients pour les arcs internes à chaque AS.

4. Résultats

Le jeu de données utilisé est issu d'historiques de tables de routage BGP proposés par les projets RIS¹⁴ [NCC] et Routeviews [UNI]. Nous avons récupéré une table de routage BGP chaque jour du 01/01/2005 au 30/04/2005. Après filtrage des AS privés, des préfixes locaux, des préfixes n'apparaissant pas plus de 90 % du temps,

14. Routing Information Service

Type du sommet	G_p	G_{pe}
AS	21021	63063
Agrégats	18640	18640
Préfixes	187970	187970
Total ($ V_p $)	227631	269673

Figure 2 – Statistiques sur les sommets des graphes G_p et G_{pe}

#bonds	couples O/D	$m = r$	$(l, s)_m = (l, s)_r$	$s_m = s_r$	$l_m > l_r \ \& \ s_m > s_r$
2	8713613	94.11 %	94.11 %	94.11 %	5.89 %
3	9039406	43.52 %	46.14 %	47.36 %	52.13 %
4	9649311	18.03 %	21.57 %	31.45 %	64.01 %
5	9937983	13.27 %	17.02 %	29.81 %	62.05 %
6	10018209	12.65 %	16.64 %	29.76 %	61.48 %
...
12	10037451	12.57 %	16.52 %	21.2 %	61.36 %

Figure 3 – Statistiques d’inférence des chemins d’AS vers préfixes de taille maximum *bond* (l : local pref, s : taille de l’AS path, m : chemin issu du modèle, r : chemin réel)

des routes provenant d’AS sources ne fournissant qu’un faible nombre de préfixes, on ne garde qu’un seul AS PATH par couple (AS source, préfixe). Les routages BGP étant très stables, il existe dans la plupart des cas un unique chemin vérifiant un maximum de jours d’observations. Pour l’inférence des accords d’interconnexion, nous avons utilisé l’algorithme CSP présenté dans [MIC 06]¹⁵. Rappelons que dans notre modèle de graphe, deux AS interconnectés sont reliés par des arcs doublement orientés. Le graphe G_p comporte donc 93234 arcs représentant 46617 accords d’interconnexion.

Du jeu de données réelles nous avons extrait les couples origine-destination (AS-préfixe) en tronquant la partie gauche de l’AS_PATH à une taille maximale donnée, puis nous avons comparé le chemin calculé aux chemins réels associés. La figure 4 montre le pourcentage de chemins concordants avec l’observation des tables BGP.

Plus le chemin observé est long, plus notre modèle trouve un chemin avec un local pref potentiel plus élevé. Pour les chemins de taille 2, il y a 6% des couples origine destination où l’AS source effectue un détour plus long pour joindre le préfixe au lieu de joindre le sommet agrégat du préfixe. La précision de 12% obtenue pour les chemins de taille maximale (12), est expliquée par des valeurs de local pref inadéquates. D’autre part, on notera que 75% des chemins terminent par le bon AS, 78% des chemins inférés ont un *LOCAL_PREF* strictement plus élevé tandis que 0.2% des

15. Notons que d’autres algorithmes ont été développés pour tenter d’inférer ces accords [GIU 04],[GAO 01], [DIM 06],[SUB 02].

routes inférées ont des *LOCAL_PREF* strictement plus faibles. Pour que les résultats soient plus précis, il faudrait avoir une meilleure estimation des *LOCAL_PREF*.

5. Conclusion et perspectives

Nous avons vu une manière de modéliser l'Internet à la granularité AS en tenant compte des préfixes. Nous avons ensuite, à l'aide d'une structure de poids et d'une transformation de graphe, proposé une méthode pour déterminer les plus courts chemins économiquement valides découlant de l'inférence des accords d'interconnexion entre AS. Notons que la détermination de k meilleurs chemins BGP entre un AS et un préfixe peut se faire en utilisant le meilleur chemin de chaque voisin de l'AS (ou même les k meilleurs de chaque voisin). Inférer de meilleures valeurs *LOCAL_PREF* pourrait permettre une simulation de l'Internet à la granularité AS plus proche de la réalité. Nous tenons finalement à remercier Benoît Besset pour son aide précieuse.

6. Bibliographie

- [DIM 06] DIMITROPOULOS X., KRIOUKOV D., FOMENKOV M., HUFFAKER B., HYUN Y., KC CLAFFY, RILEY G., « AS Relationships : Inference and Validation », *in submission to ACM Computer Communication Review*, 2006.
- [GAO 01] GAO L., « On inferring autonomous system relationships in the Internet », *IEEE/ACM Trans. Netw. vol. 9, no. 6, pp. 733-745*, 2001.
- [GIU 04] GIUSEPPE DI BATTISTA T. E. . A., « DELIS-TR-0001 - Computing the Types of the Relationships between Autonomous Systems », techreport n° 0001, 2004, DELIS.
- [MEU 07] MEULLE M., « Inférence des accords économiques et des politiques de routage dans l'Internet », PhD thesis, 2007.
- [M.G 02] M. GONDRAN M. M., « Graphes, dioïdes, et semi anneaux », , 2002, p. 127–186, Tec & doc.
- [MIC 06] MICKAEL MEULLE QUANG NGUYEN J.-K. H., « Formulation CSP et approches heuristiques pour l'inférence des accords d'interconnexion dans l'Internet », *Recherche Opérationnelle et Aide à la DEcision Française (ROADEF'02) proceedings*, 2006.
- [M.M 04] M. MEULLE Q. N., « Un modèle de graphe pour le routage BGP », *Algotel*, 2004.
- [NCC] NCC R., « Routing Information Service Raw Data, <http://data.ris.ripe.net/> ».
- [QUO 05] QUOITIN B., PELSSER C., BONAVENTURE O., UHLIG S., « A performance evaluation of BGP-based traffic engineering », *Int. J. Netw. Manag.*, vol. 15, n° 3, 2005, p. 177–191, John Wiley & Sons, Inc.
- [SUB 02] SUBRAMANIAN L., AGARWAL S., REXFORD J., KATZ R., « Characterizing the Internet hierarchy from multiple vantage points », *INFOCOM'02, IEEE, Volume 2, 23-27*, 2002.
- [UNI] UNIVERSITY O., « Route Views Project archives, <http://archive.routeviews.org/> ».
- [WAN 03] WANG F., GAO L., « On inferring and characterizing internet routing policies », *IMC '03 : Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, New York, NY, USA, 2003, ACM Press, p. 15–26.